



## PDF hosted at the Radboud Repository of the Radboud University Nijmegen

The version of the following full text has not yet been defined or was untraceable and may differ from the publisher's version.

For additional information about this publication click this link.

<http://hdl.handle.net/2066/18715>

Please be advised that this information was generated on 2017-12-05 and may be subject to change.

DEPARTMENT OF MATHEMATICS  
UNIVERSITY OF NIJMEGEN The Netherlands

**Pure equilibrium strategies for  
Stochastic Games via Potential Functions**

**Jos AM Potters, TES Raghavan and Stef H Tijs**

Addresses:

Jos AM Potters, Mathematical Institute KUN, Toernooiveld 1,  
6525 ED Nijmegen, The Netherlands.      e-mail: [potters@sci.kun.nl](mailto:potters@sci.kun.nl)

TES Raghavan, UIC, Department of MSCS, 851 S. Morgan  
Chicago, Illinois 60607-7045, USA      e-mail: [ter@uic.edu](mailto:ter@uic.edu)

Stef H Tijs, Department of Econometrics, P.O. Box 90153,  
5000 LE Tilburg, The Netherlands      e-mail: [S.H.Tijs@kub.nl](mailto:S.H.Tijs@kub.nl)

## Pure equilibrium strategies for Stochastic Games via Potential Functions

**Abstract** Strategic games with a potential function have quite often equilibria in pure strategies (Monderer and Shapley (1996)). This is also true for stochastic games but the existence of a potential function is mostly hard to prove. For some classes of stochastic games with an additional structure an equilibrium can be found by solving one or a finite number of finite strategic games. We call these games *auxiliary games*. In the paper we investigate if we can derive the existence of equilibria in pure stationary strategies from the fact that the auxiliary games allow for a potential function. We will do this for zero-sum two-person discounted stochastic games and non-zero-sum discounted stochastic games with additive reward functions and additive transitions (Raghavan et al. (1985)) or with separable rewards and state independent transitions (Parthasarathy et al. (1984)).

Key words: Stochastic games, strategic game with a potential function.

JEL-classification C72, C73

### Introduction.

A potential game is a strategic game that allows a (potential) function on the set of strategy profiles such that potential differences under a *unilateral deviation* are the same as the differences in payoff for the deviating player. Therefore, the potential function can be used by deviating players as a leading principle in the sense that a profitable deviation increases the value of the potential function and, conversely, if the potential value increases under unilateral deviation, the deviation is profitable for the deviating player. This has two important consequences: an improvement process (a series of profitable unilateral deviations by various players) cannot cycle and a maximum of the potential function (if it exists) is automatically a Nash equilibrium. (Monderer and Shapley (1996)). For finite strategic games this means that a potential game has a *pure* Nash equilibrium and every (neatly defined) improvement process is finite.

In the theory of stochastic games one finds several subclasses that can be solved by considering and solving one or more—what we will call—*finite auxiliary strategic games*. The solution of the auxiliary game(s) leads to a solution of the stochastic

game. Examples are the zero-sum games occurring in the Shapley equations for zero-sum two-person stochastic games (Shapley (1953)), non-zero-sum stochastic game with an ARAT-structure (Additive Rewards and Additive Transitions, Raghavan et al. (1985)) or a SER-SIT-structure (Separable Rewards and State Independent Transitions, Parthasarathy et al. (1984)).

In this paper we propose to investigate these classes of stochastic games and to see whether the existence of *pure stationary equilibria* (*saddle points* in the zero-sum case) can be derived from the existence of a potential function in the auxiliary game(s). The idea of using a pure stationary strategy is less problematic than the use of mixed actions in some states of the stochastic game.

Section 1 introduces the fundamental concepts in the theory of potential games (subsection 1.1) and the theory of  $\beta$ -discounted two-person stochastic games (subsection 1.2). Some of the results of the paper can be extended to  $n$ -person or undiscounted stochastic games but the paper focusses on  $\beta$ -discounted two-person games.

Section 2 discusses the zero-sum case and shows that a *zero-sum*  $\beta$ -discounted two-person stochastic game has a saddle point if the auxiliary games occurring in the Shapley equation (Shapley (1953)) have a potential function. *Zero-sum ARAT stochastic games* have these property but they form only a subset as we will show by an example.

Section 3 deals with non-zero-sum stochastic games. We consider the class of two-person stochastic games in which one player has additive rewards and the other player controls the transitions. These games are proved to have pure stationary equilibria. Another class we will investigate is formed by SER-SIT games (Parthasarathy et al. (1984)). Here we need an additional condition to find a potential function in the auxiliary game introduced by Parthasarathy et al. By an example we show that the (or at least an) additional condition is needed.

## 1. The models and the tools.

Subsection 1.1 contains the basic definitions in the theory of strategic games with a potential function and Section 1.2 recalls the basic facts about stochastic games (cf. the book of Filar and Vrieze (1996) for a comprehensive study of the subject).

### 1.1 Strategic Games with a Potential Function.

Potential games are introduced in Monderer and Shapley (1996).

An  $n$ -person strategic game  $\langle A_1, \dots, A_n, u_1, \dots, u_n \rangle$  with (finite) action space  $A_i$  and

utility function  $u_i: A = \prod_{i=1}^n A_i \rightarrow \mathbb{R}$  for each player  $i \in N$  is called a *potential game* if there is a *potential function*  $F: A \rightarrow \mathbb{R}$  such that, for all  $i \in N$ , for all strategy profiles  $(a_i)_{i \in N} \in A$  and every alternative action  $b_i \in A_i$ , we have  $u_i(b_i, a_{-i}) - u_i(a) = F(b_i, a_{-i}) - F(a)$ .

Here  $a_{-i} = (a_1, \dots, a_{i-1}, a_{i+1}, \dots, a_n)$  and  $(b_i, a_{-i}) = (a_1, \dots, a_{i-1}, b_i, a_{i+1}, \dots, a_n)$

Interesting properties of such a potential game are the existence of pure Nash equilibria (each element of  $\arg\max(F)$  is a Nash equilibrium) and the so called finite improvement property: starting from any strategy profile a Nash equilibrium is reached after a finite number of unilateral improvements by various players.

In the next proposition we give two characterizations for two-person zero-sum potential games. One characterization says that, for each  $2 \times 2$ -subgame, the sum of the payoffs in the diagonal and the anti-diagonal are the same. The other characterization says that the utility function for each player is the sum of a part only dependent on the player's own action, and a part only dependent on the action of his opponent.

**Proposition 1.** *Given a strategic zero-sum game  $\langle A_1, A_2, u_1, u_2 \rangle$  ( $u_1 + u_2 = 0$ ) the following assertions are equivalent:*

(i)  $\langle A_1, A_2, u_1, u_2 \rangle$  is a potential game

(ii) (diagonal property) For all  $a_1, b_1 \in A_1$  and  $a_2, b_2 \in A_2$  we have

$$u_1(a_1, a_2) + u_1(b_1, b_2) = u_1(a_1, b_2) + u_1(b_1, a_2)$$

(iii) (separation property) There are functions  $g_1: A_1 \rightarrow \mathbb{R}$  and  $g_2: A_2 \rightarrow \mathbb{R}$  such that

$$u_1(a_1, a_2) = g_1(a_1) + g_2(a_2) \text{ for all } (a_1, a_2) \in A_1 \times A_2.$$

**Proof:** (a) Monderer and Shapley (1996, Theorem 2.8) proved that  $\langle A_1, A_2, u_1, u_2 \rangle$  is a potential game if and only if, for all  $a_i, b_i \in A_i$ , we have

$$\begin{aligned} & (u_1(a_1, a_2) - u_1(b_1, a_2)) + (u_2(b_1, a_2) - u_2(b_1, b_2)) + \\ & (u_1(b_1, b_2) - u_1(a_1, b_2)) + (u_2(a_1, b_2) - u_2(a_1, a_2)) = 0. \end{aligned}$$

Substituting  $u_2 = -u_1$  gives the equality

$$2(u_1(a_1, a_2) - u_1(b_1, a_2)) - 2u_1(a_1, b_2) + 2u_1(b_1, b_2) = 0. \text{ This is assertion (ii).}$$

(b) We assume the diagonal property. Take two points of reference  $a_1^\circ \in A_1$  and  $a_2^\circ \in A_2$  and define the functions  $g_1$  and  $g_2$  by  $g_1(a_1) := u_1(a_1, a_2^\circ)$  and  $g_2(a_2) := u_1(a_1^\circ, a_2) - u_1(a_1^\circ, a_2^\circ)$ . Then,

$$\begin{aligned} u_1(a_1, a_2) &= (u_1(a_1, a_2) - u_1(a_1^\circ, a_2)) + (u_1(a_1^\circ, a_2) - u_1(a_1^\circ, a_2^\circ)) + u_1(a_1^\circ, a_2^\circ) = \\ &= (u_1(a_1, a_2) + u_1(a_1^\circ, a_2^\circ) - u_1(a_1^\circ, a_2)) + g_2(a_2) = g_1(a_1) + g_2(a_2) \end{aligned}$$

because of property (ii).

(c) If  $u_1$  is separable i.e.,  $u_1(a_1, a_2) = g_1(a_1) + g_2(a_2)$  for certain functions  $g_i: A_i \rightarrow \mathbb{R}$  ( $i = 1, 2$ ), the diagonal property is easy to prove.  $\triangleleft$

**Remark:** For non-zero-sum strategic games the diagonal property and the separability property are not needed for the existence of a potential function, as the bimatrix game with the following payoff matrices and potential function shows:

$$(X, Y) = \begin{bmatrix} (1, 5) & (1, 0) \\ (3, 10) & (4, 6) \end{bmatrix} \quad \text{and} \quad F = \begin{bmatrix} 5 & 0 \\ 7 & 3 \end{bmatrix}.$$

Neither of the two payoff matrices has the diagonal property or the separability property.

For bimatrix games the equivalence of the diagonal property and the separability property remains true and these properties are still sufficient for the existence of a potential function. Necessary is that the difference of the payoff matrices satisfies the diagonal property.

## 1.2 Stochastic games with pure Nash equilibria.

A *two-person stochastic game* is determined by the following data:

- There is a finite set  $S$  of *states*. Elements of  $S$  are denoted by  $s, t, \dots$
- There are two players, I and II.
- In each state  $s \in S$  each of the players has a finite set of *actions*:  $A_s^1$  and  $A_s^2$ .
- There are two functions  $r^1$  and  $r^2$  (the *reward functions*) defined on the set  $T := \{(s, a^1, a^2) : s \in S, a^1 \in A_s^1 \text{ and } a^2 \in A_s^2\}$ .
- There is a map  $p: T \rightarrow \Delta(S)$  (the set of probability vectors on  $S$ ). We write  $p(t | s, a^1, a^2)$  for the  $t$ -th coordinate of  $p(s, a^1, a^2)$ . The map  $p$  gives the *Markov transition probabilities*.
- The game is played in a finite or countable sequence of *stages*. In each stage the players know what happened in the preceding stages and in particular the present state  $s$  and choose independently actions  $a^1$  and  $a^2$  in their own action spaces  $A_s^1$  and  $A_s^2$ . Their action choice can be based on all information they have at the moment they make their decision: the history of the game. An history has the following form:

$$(s_0, a_0^1, a_0^2, s_1, a_1^1, a_1^2, \dots, a_{n-1}^1, a_{n-1}^2, s_n).$$

The action choice results in an immediate reward to each of the players,  $r^k(s, a^1, a^2)$  is the reward to player  $k$  and the next stage the process is in state  $t$  with probability  $p(t | s, a^1, a^2)$ . The history (past states and past actions) of the game is common knowledge between the players.

By playing the game each of the players obtains a payoff every period and a stream of payoffs  $\{r^k(s_n, a_n^1, a_n^2)\}_{n \in \mathbb{N}_0}$  results.

We assume that both players appreciate such a flow of rewards according to the utility function  $U^k(\{r^k(s_n, a_n^1, a_n^2)\}_{n \in \mathbb{N}_0}) := \sum_{n \in \mathbb{N}_0} \beta^n r^k(s_n, a_n^1, a_n^2)$ .

Here  $\beta \in (0, 1)$  is a fixed *discount factor*. If there are only finitely many stages (a stochastic game with finite horizon) the summation is taken over the finitely many stages.

A stochastic game is called *zero-sum* if  $r^1 + r^2 = 0$  on  $T$ .

A *Markov component for player  $k$*  is a map  $f^k: S \rightarrow \bigcup_{s \in S} \Delta(A_s^k)$  with the property  $f^k(s) \in \Delta(A_s^k)$ . So, a Markov component of player  $k$  determines a mixed action in  $A_s^k$  for each state  $s \in S$  and the action choice is only dependent on the present state, the last entry of a history.  $MC_k$  denotes the (finite) set of Markov components of player  $k$ . A *Markov strategy* consists of a sequence of Markov components  $\{f_n^k\}_{n \in \mathbb{N}_0}$ , one for each stage  $n$ . If  $f_n^k$  is the same for every stage, the strategy is called a *stationary (Markov) strategy*. If  $f_n^k(s)$  is a pure action for every stage  $n$  and every state  $s$ , the (Markov or stationary) strategy is called a *pure strategy*. We denote Markov strategies for player  $k$  by  $\bar{f}^k$  and the stationary strategy with  $f_n^k = f^k$  for all  $n$  by  $(f^k)^\infty$ . If Markov strategies (or more general history dependent strategies)  $\bar{f}^1$  and  $\bar{f}^2$  are chosen, a stochastic process determines the probabilities to be in stage  $s_n$  at time  $n$  (as a function of the initial state  $s_0$ ) and the stream of rewards  $\{r^k(s_n, f_n^1(s_n), f_n^2(s_n))\}_{n \in \mathbb{N}_0}$  for both players. We define the payoff functions by

$$V_\beta^k(\bar{f}^1, \bar{f}^2)_{s_0} := \sum_{n \in \mathbb{N}_0} \beta^n r^k(s_n, f_n^1(s_n), f_n^2(s_n)) \quad (k = 1, 2)$$

for all initial states  $s_0 \in S$ .

Then we have a strategic game (dependent on the set of strategies we allow). Nash equilibria of this game are called the optimal (or equilibrium) solutions of the  $\beta$ -discounted stochastic game.

For  $\beta$ -discounted two-person zero-sum stochastic games we have the following fundamental result of Shapley (1953).

- (i) Both players have a stationary optimal strategy. So, a  $\beta$ -discounted two-person zero-sum stochastic games can be solved in stationary strategies.
- (ii) To find optimal stationary strategies  $(f^1)^\infty$  and  $(f^2)^\infty$  and the value function  $v_\beta: S \rightarrow \mathbb{R}$  defined by  $v_\beta(s) := V_\beta^1((f^1)^\infty, (f^2)^\infty)$ , a collection of *auxiliary* zero-sum games is considered.

For every state  $s$  and every vector  $x \in \mathbb{R}^S$  we define the finite zero-sum game  $\Gamma(s, x)$  with action space  $A_s^1$  and  $A_s^2$  and payoff matrix

$$[r^1(s, a^1, a^2) + \beta \sum_{t \in S} p(t | s, a^1, a^2) x_t]_{a^1 \in A_1^1, a^2 \in A_2^2}.$$

If a vector  $x \in \mathbb{R}^S$  satisfies the *Shapley equations*

$$x_s = \text{value } (\Gamma(s, x)) \text{ for every } s \in S,$$

then  $x = v_\beta$  and optimal strategy profiles  $(f^1(s), f^2(s))$  in the auxiliary games  $\Gamma(s, x = v_\beta)$  determine a stationary optimal solution  $((f^1)^\infty, (f^2)^\infty)$  of the stochastic game.

Also, non-zero-sum  $\beta$ -discounted two-person stochastic games are solvable in stationary strategies. This result was proved by Fink (1964) and Takahashi (1964). Both authors used a fixed point argument.

If in a stochastic game the Markov strategies  $\bar{f}^1 = \{f_n^1\}_{n \in \mathbb{N}_0}$  and  $\bar{f}^2 = \{f_n^2\}_{n \in \mathbb{N}_0}$  are applied and  $x_0 \in \Delta(S)$  gives the initial probability distribution over the states, the probability distribution  $x_n$  over the states at time  $n$  is given by the recursive formula

$$x_n(t) = \sum_{s \in S} x_{n-1}(s) p(t | s, f_{n-1}^1(s), f_{n-1}^2(s)) \text{ for all } t \in S.$$

If we introduce the  $S \times S$ -stochastic matrix

$P(f^1, f^2)$  by  $P(f^1, f^2)_{s,t} := p(t | s, f^1(s), f^2(s))$  for every pair of Markov components  $f^1, f^2$ , the latter equation can be written as

$$x_n = x_{n-1} P(f_{n-1}^1, f_{n-1}^2) \text{ for } n = 1, 2, \dots$$

The expected payoff in stage  $n$ , when  $(\bar{f}^1, \bar{f}^2)$  is applied and the original distribution over the states is  $x_0$ , equals

$$\begin{aligned} \sum_{s \in S} x_n(s) r(s, f_n^1(s), f_n^2(s)) &= \\ &= \sum_{s \in S} (x_0 P(f_0^1, f_0^2) \cdots P(f_{n-1}^1, f_{n-1}^2))_s r(s, f_n^1(s), f_n^2(s)). \end{aligned}$$

If we define the  $S$ -vector  $R^k(f^1, f^2) \in \mathbb{R}^S$  by  $R^k(f^1, f^2)_s := r^k(s, f^1(s), f^2(s))$  for every pair of Markov components  $(f^1, f^2)$ , we find for the expected payoff in period  $n$

$$\begin{aligned} \sum_{s \in S} x_0(s) P(f_0^1, f_0^2) \cdots P(f_{n-1}^1, f_{n-1}^2) R^k(f_n^1, f_n^2)_s &= \\ \langle x_0, P(f_0^1, f_0^2) \cdots P(f_{n-1}^1, f_{n-1}^2) R^k(f_n^1, f_n^2) \rangle. \end{aligned}$$

For stationary strategies the  $\beta$ -discounted payoff vectors  $V_\beta^k((f^1)^\infty, (f^2)^\infty)$  ( $k = 1, 2$ ) equals

$$\begin{aligned} V_\beta^k((f^1)^\infty, (f^2)^\infty) &= R^k(f^1, f^2) + \beta P(f^1, f^2) R^k(f^1, f^2) + \cdots + \\ &+ \beta^{n-1} P(f^1, f^2)^{n-1} R^k(f^1, f^2) + \cdots = [I - \beta P(f^1, f^2)]^{-1} R^k(f^1, f^2). \end{aligned}$$

Notice that the matrix  $[I - \beta P(f^1, f^2)]^{-1}$  is a nonnegative matrix for every pair of



Markov components  $(f^1, f^2)$ .

The main topic of this paper is to find conditions under which two-person  $\beta$ -discounted stochastic games have pure stationary equilibria. In the literature there are several special classes of stochastic games that can be solved by using ‘auxiliary games’ of various nature. We will investigate for which classes the existence of pure stationary strategies can be derived from the fact that the auxiliary game has a potential function.

We will repeat the definition of the classes we will consider.

ARAT-games (Additive Rewards and Additive Transitions)

A two-person stochastic game is an *ARAT-game* if, for all  $s \in S$ , all  $a^1 \in A_s^1$  and all  $a^2 \in A_s^2$ ,

$$\begin{aligned} r^1(s, a^1, a^2) &= r^{11}(s, a^1) + r^{12}(s, a^2), \\ r^2(s, a^1, a^2) &= r^{21}(s, a^1) + r^{22}(s, a^2), \\ p(t | s, a^1, a^2) &= p^1(t | s, a^1) + p^2(t | s, a^2) \end{aligned}$$

for some functions  $r^{ij}$  and  $p^j$ . So, the reward functions as well as the transition probabilities are sums of two functions, one only dependent on the state and the action of player I, the other only dependent on the state and the action of player II. If  $p^1$  or  $p^2$  is identically zero, we talk about *single control games*. If, for every element  $s \in S$  one of the transition functions  $p^1(s, a^1, a^2)$  or  $p^2(s, a^1, a^2)$  vanishes, we call the stochastic game a stochastic game with *switching control*.

SER-SIT-games (Separable Rewards and State Independent Transitions)

In a SER-SIT stochastic game the actions spaces are the same in each state:

$$A_s^k = A_t^k (= A^k) \text{ for } s, t \in S \text{ and } k = 1, 2.$$

A stochastic game is a *SER-SIT-game* if, for all states  $s \in S$ , all  $a^1 \in A^1$  and all  $a^2 \in A^2$ ,

$$\begin{aligned} r^1(s, a^1, a^2) &= r^{10}(s) + r^{11}(a^1, a^2) \\ r^2(s, a^1, a^2) &= r^{20}(s) + r^{21}(a^1, a^2) \\ p(t | s, a^1, a^2) &= p(t | a^1, a^2) \text{ is not dependent on } s. \end{aligned}$$

So, the reward function is the sum of two parts, one only dependent on the present state and the other only dependent on the action profile. The transitions are not dependent on the present state.

If  $p(t | s, a^1, a^2) = p(t | s)$  is not dependent on the actions, we call the stochastic game a stochastic game with *action independent transitions* (AIT).

## 2. Zero-sum stochastic games with pure optimal stationary strategies.

In this section we prove that two-person zero-sum  $\beta$ -discounted stochastic games have a *pure* optimal stationary strategy if each of the auxiliary games  $\Gamma(s, v_\beta)$  ( $s \in S$ ) has a potential function. If all auxiliary games  $\Gamma(s, x)$  ( $s \in S, x \in \mathbb{R}^S$ ) have a potential function, then the stochastic game is an ARAT-game. We provide an example showing that not every stochastic game of the first category is an ARAT-game.

**Lemma 2.** *If the auxiliary games  $\{\Gamma(s, v_\beta)\}_{s \in S}$  of a two-person zero-sum  $\beta$ -discounted stochastic game have a potential function, then the stochastic game has an optimal stationary strategy in pure strategies.*

**Proof:** By Monderer and Shapley (1996) every auxiliary game has a pure equilibrium and by Shapley (1953) these pure strategies form an optimal stationary strategy in the stochastic game.  $\triangleleft$

In the second lemma we prove that Lemma 2 can be applied to zero-sum ARAT-games.

**Lemma 3.** *If a two-person zero-sum game is an ARAT-game, all auxiliary games  $\Gamma(s, x)$  have a potential function and, conversely, if all auxiliary games  $\Gamma(s, x)$  have a potential function, the stochastic game is an ARAT-game.*

**Proof:** The payoff in any game  $\Gamma(s, x)$  can be written as

$$r^1(s, a^1, a^2) + \beta \sum_{t \in S} p(t | s, a^1, a^2) x_t = [r^{11}(s, a^1) + \beta \sum_{t \in S} p^1(t | s, a^1) x_t] + [r^{12}(s, a^2) + \beta \sum_{t \in S} p^2(t | s, a^2) x_t].$$

So, the zero-sum game  $\Gamma(s, x)$  satisfies the separability property and, therefore, it admits a potential function by Proposition 1.

Conversely, if  $\Gamma(s, x)$  for all  $s \in S$  and all  $x \in \mathbb{R}^S$ , one can take  $x = 0$  and find by Proposition 1, that the rewards are additive:  $r^1(s, a^1, a^2) = r^{11}(s, a^1) + r^{12}(s, a^2)$ .

If we take any  $s, t \in S$  and  $x = e_t$ , the separation condition for the auxiliary game  $\Gamma(s, e_t)$  gives  $r^{11}(s, a^1) + r^{12}(s, a^2) + \beta p(t | s, a^1, a^2) = \bar{r}^{11}(s, a^1) + \bar{r}^{12}(s, a^2)$  for certain functions  $\bar{r}^{11}$  and  $\bar{r}^{12}$ .

From this equation we can find the components

$$p^k(t | s, a^k) = \frac{\bar{r}^{1k}(s, a^k) - r^{1k}(s, a^k)}{\beta}. \quad \triangleleft$$

**Corollary.** *Two-person zero-sum ARAT-games have a pure optimal stationary strategy.*

The next example shows that the ARAT-games do not exhaust the class of stochastic games where Lemma 2 applies.

**Example:** Let  $S = \{s_0, s_1, s_2, s_3\}$ . The states  $s_i \neq s_0$  are absorbing states: both

players have one action and the reward is  $r(s_i) = u_i$  for  $i = 1, 2, 3$ . In state  $s_0$  both players have two actions and the rewards and transitions are given by

$$r^1(s_0) = \begin{bmatrix} -u_1 & -u_2 \\ -u_3 & 0 \end{bmatrix} \quad \text{and} \quad p(s_0) = \begin{bmatrix} \rightarrow s_1 & \rightarrow s_2 \\ \rightarrow s_3 & \rightarrow s_0 \end{bmatrix}.$$

Then for the state  $s_i$  ( $i \geq 1$ ) we have  $v_\beta(s_i) = \frac{u_i}{1-\beta}$  and for state  $s_0$  the auxiliary game is

$$\Gamma(s_0, v_\beta) : \begin{bmatrix} u_1(-1 + \frac{\beta}{1-\beta}) & u_2(-1 + \frac{\beta}{1-\beta}) \\ u_3(-1 + \frac{\beta}{1-\beta}) & \beta v_\beta(s_0) \end{bmatrix}.$$

Then  $v_\beta(s_0) = 0$  solves the Shapley equations if

$$u_3(-1 + \frac{\beta}{1-\beta}) > 0 \text{ and } u_2(-1 + \frac{\beta}{1-\beta}) < 0$$

If we take  $u_1 = u_2 + u_3$  we have the diagonal property. If  $\beta > 0.5$  we must have  $u_3 > 0$ ,  $u_2 < 0$  and  $u_1 = u_2 + u_3$ . Clearly, the rewards are additive but the transitions are not

### 3. Non-zero-sum stochastic games with equilibria in pure strategies.

We start with two rather simple classes of non-zero-sum stochastic games with pure Nash equilibria. The first class is the class of *coordination games*. A  $n$ -person stochastic game is a *coordination game* if all players have the *same reward function*. The second class is the class of stochastic games with action independent transitions in which all  $n$ -person games  $a = (a_1, \dots, a_n) \mapsto (r^1(s, a), \dots, r^n(s, a))$  have a potential function.

#### Theorem 4.

(i) *If in an  $n$ -person stochastic game the reward functions are the same for all players, then the stochastic game has a Nash equilibrium in pure strategies.*

(ii) *If an  $n$ -person stochastic game has action independent transitions (AIT) and for each state  $s \in S$  the finite  $n$ -person game with strategy sets  $\{A_s^k\}_{k=1, \dots, n}$  and payoff functions  $a \mapsto \{r^k(s, a)\}_{k=1, \dots, n}$  has a potential function, then the stochastic game has a Nash equilibrium in pure strategies.*

**Proof:** (i) We prove that a Nash equilibrium is obtained by solving the following dynamical programming problem. The set of states  $\bar{S}$  is the same as in the stochastic game. The actions in state  $s \in S$  is  $\bar{A}_s := \prod_{i=1}^n A_s^i$ . Finally the reward functions and transition functions are also the same as in the stochastic game (although the interpretation is slightly different):  $\bar{r} = r = r^i$  for all players  $i$  and  $\bar{p} = p$ . It is well known (see Blackwell (1962)) that dynamic programming problems have optimal

stationary strategies in pure actions. Let  $\bar{f} : s \in S \mapsto \bar{f}(s) \in \bar{A}_s$  be such an optimal strategy. If each player  $i$  chooses the  $i$ -th component  $f_i$  of  $\bar{f}$ , this is a Nash equilibrium in the stochastic game. If any player deviates from his strategy in any stage of the game, his expected payoff cannot increase as the same deviation in the dynamical program would give the same increase of expected payoff.

(ii) If the players have no influence on the transitions, they will make the best out of the state they are. By the result of Shapley and Monderer there is a pure Nash equilibrium in each *auxiliary game*. That is, for each state  $s \in S$  there is a Nash equilibrium  $(f^1(s), \dots, f^n(s)) \in \prod_{i=1}^n A_s^i$  of the stage game. This defines a Markov component  $f^i$  for each player  $i$ . Deviation of any player will not increase the payoff at any time in any state and has no influence on later expected payoffs. No player can gain by deviating at any time in any state.  $\triangleleft$

For non-zero-sum stochastic games the ARAT-structure is not sufficient for the existence of a pure stationary equilibrium (see Raghavan et al.(1985) and Raghavan Thuijsman (1997) for an example). Moreover, the auxiliary games as considered in the case of zero-sum games make little sense. If, however, the rewards for one player are *additive* and the transitions are *single controlled* by the other player, there are pure stationary equilibria.

**Theorem 5.** *If in a non-zero-sum two-person stochastic game player I has an additive reward function and the transitions are controlled by (the actions of) player II, then there is a pure stationary equilibrium.*

**Proof:** The additivity of the reward function of player I means that

$$r^1(s, a^1, a^2) = r^{11}(s, a^1) + r^{12}(s, a^2).$$

The single control says that  $p(t | s, a^1, a^2) = p^2(t | s, a^2)$ .

As player I has no influence on the transitions it seems natural to assume that, in every state he will maximize the part of the reward  $r^{11}(s, a^1)$  under his control.

So, we define a pure Markov component  $f_0^1$  with the property that

$$r^{11}(s, f_0^1(s)) = \max_{a^1 \in A_s^1} r^{11}(s, a^1).$$

If we fix  $f_0^1$  player II will consider the dynamic decision problem with reward function  $\bar{r} : T^2 := \{(s, a^2) : s \in S, a^2 \in A_s^2\} \rightarrow \mathbb{R}$  defined by  $\bar{r}(s, a^2) := r^2(s, f_0^1(s), a^2)$  and transition probabilities  $\bar{p} = p^2 : T^2 \rightarrow \Delta(S)$  as before.

Dynamical Programming problems can be solved in pure stationary strategies. One solves first the *Bellman equations*:

$$x \in \mathbb{R}^S : \quad x_s = \max_{a^2 \in A_s^2} [\bar{r}(s, a^2) + \beta \sum_{t \in S} \bar{p}(t | s, a^2) x_t]$$

and takes  $f_0^2(s) \in A_s^2$  such that

$$f_0^2(s) \in \arg \max \{a^2 \mapsto [\bar{r}(s, a^2) + \beta \sum_{t \in S} \bar{p}(t | s, a^2) x_t]\}.$$

Then  $(f_0^2)^\infty$  is a pure stationary strategy and a best response to  $(f_0^1)^\infty$ , because it is an optimal strategy in the dynamic program.

The pure stationary strategy  $(f_0^1)^\infty$  is a best response to every stationary strategy of player II. This follows from the formula  $V_\beta^1(f_0^1, f^2) = [I - \beta P(f^2)]^{-1} (R^{11}(f_0^1) + R^{12}(f^2))$ . If player I deviates from  $f_0^1$  to  $f^1$  all components of  $R^{11}(f^1)$  are less or equal to the components of  $R^{11}(f_0^1)$  and as the matrix  $[I - \beta P(f^2)]^{-1} \geq 0$  also

$$[I - \beta P(f^2)]^{-1} (R^{11}(f_0^1) - R^{11}(f^1)) \geq 0.$$

The term  $[I - \beta P(f^2)]^{-1} (R^{12}(f^2))$  does not change by a deviation of player I and  $(f_0^1)^\infty$  is a best response to every stationary strategy of player II.  $\triangleleft$

**Remark:** In Nowak and Raghavan (1993) it is proved that, if the transitions are single control, a Nash equilibrium can be found by solving the following *auxiliary* bimatrix game. The strategy space of each player consists of the finite set of all Markov components  $\{f_k^1\}_{k \in MC^1}$  and  $\{f_l^2\}_{l \in MC^2}$ .

If player I chooses the Markov component  $f_k^1$  and player II chooses the Markov component  $f_l^2$ , the payoff to player I is  $\sum_{s \in S} R^1(f_k^1, f_l^2)_s$  and the payoff to player II is  $\sum_{s \in S} V_\beta^2(f_k^1, f_l^2)_s$ . If  $\{\xi_k\}_{k \in MC^1}$  and  $\{\eta_l\}_{l \in MC^2}$  form a Nash equilibrium of the auxiliary game, the stationary strategies  $f_*^1 = \sum_{k \in MC^1} \xi_k f_k^1$  and  $f_*^2 = \sum_{l \in MC^2} \eta_l f_l^2$  defines an equilibrium in stationary strategies for the  $\beta$ -discounted stochastic game. Therefore, the stochastic game has a pure stationary strategy, if the auxiliary bimatrix game has a pure Nash equilibrium.

If we have a SER-SIT-stochastic game and the bimatrix game  $[r^k(s, a^1, a^2)]_{a^1 \in A^1, a^2 \in A^2}$  is a potential game for every state  $s \in S$ , the  $\beta$ -discounted stochastic game need not have a pure stationary equilibrium. We need an additional condition.

**Theorem 6.** *If in a stochastic game with separable reward functions and state independent transitions*

- (i) *the partial reward game  $[r^{k1}(s, a^1, a^2)]_{a^1 \in A^1, a^2 \in A^2}$  has a potential function and*
  - (ii) *the matrix game  $Q: (a^1, a^2) \rightarrow \langle p(a^1, a^2), r^{10} - r^{20} \rangle$  has the diagonal property,*
- then the stochastic game has an equilibrium in pure stationary strategies.*

**Proof:** Parthasarathy et al. (1984) provides a method to solve SER-SIT stochastic games. It is sufficient to solve the auxiliary bimatrix game with strategy spaces  $A^1$  and  $A^2$  and payoff matrices  $[r^{k1}(a^1, a^2) + \beta \sum_{t \in S} p(t | a^1, a^2) r^{k0}(t)]_{a^1 \in A^1, a^2 \in A^2}$ . If  $(\xi, \eta) \in \Delta(A^1) \times \Delta(A^2)$  is a Nash equilibrium of the auxiliary game, then the (even state independent) stationary strategy  $(\xi^\infty, \eta^\infty)$  is an equilibrium in the  $\beta$ -discounted SER-SIT stochastic game. If the auxiliary game has a potential function,

the stochastic game has a pure stationary equilibrium.

Let  $F_1: A^1 \times A^2$  be a potential function of the reward game  $[r^{k1}(a^1, a^2)]_{a^1 \in A^1, a^2 \in A^2}$ . We have to prove that the bimatrix game  $[\beta \sum_{t \in S} p(t | a^1, a^2) r^{k0}(t)]$  has a potential function. By Monderer and Shapley (1996) (Theorem 2.8) this is true if and only if the matrix

$Q = [\beta \sum_{t \in S} p(t | a^1, a^2) (r^{10}(t) - r^{20}(t))]_{a^1 \in A^1, a^2 \in A^2}$  has the diagonal property (see the proof of Proposition 1).  $\triangleleft$

**Example** The following example shows that the condition

$Q: (a^1, a^2) \rightarrow \langle p(a^1, a^2), r^{10} - r^{20} \rangle$  has the diagonal property, is not superfluous. Let  $S = \{s_1, s_2\}$ . The action spaces  $A^1$  and  $A^2$  consist of two actions.

$$r^{k1}(a^1, a^2) := \begin{bmatrix} (1, 0) & (0, 0) \\ (2, 2) & (0, 1) \end{bmatrix} \quad r^{k0} \begin{pmatrix} s_1 \\ s_2 \end{pmatrix} := \begin{bmatrix} (3 & -1) \\ (-3 & 1) \end{bmatrix}$$

$$p(s_1 | a^1, a^2) : p(s_2 | a^1, a^2) := \begin{bmatrix} (1 : 0) & (0 : 1) \\ (0.5 : 0.5) & (0.5 : 0.5) \end{bmatrix}.$$

A potential function  $F_1$  for  $[r^{k1}]$  and the function  $Q: (a^1, a^2) \rightarrow \langle p(a^1, a^2), r^{10} - r^{20} \rangle$  have the following values

$$F_1(a^1, a^2) = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \quad Q(a^1, a^2) = \begin{bmatrix} 4 & -4 \\ 0 & 0 \end{bmatrix} \text{ (not satisfying the diagonal property).}$$

If we compute the auxiliary bimatrix game, we get the matrices:

$$\begin{bmatrix} (1, 0) & (0, 0) \\ (2, 2) & (0, 1) \end{bmatrix} + \beta \begin{bmatrix} (3, -1) & (-3, 1) \\ (0, 0) & (0, 0) \end{bmatrix} = \begin{bmatrix} (1 + 3\beta, -\beta) & (-3\beta, \beta) \\ (2, 2) & (0, 1) \end{bmatrix}.$$

If  $\beta > \frac{1}{3}$ , the auxiliary game has only one *completely mixed* Nash equilibrium.

## References.

- Blackwell D (1962) Discounted dynamic programming. *Annals of Mathematical Statistics*, **33**, 719–726.
- Filar JA and Raghavan TES (1984) A matrix game solution of the single controller stochastic game. *Mathematics of Operations Research*, **4**, 356–362.
- Filar JA and Vrieze OJ (1996) Competitive Markov decision processes. Springer Verlag, Berlin.
- Fink AM (1964) Equilibrium in stochastic  $n$ -person game. *J. Science of the Hiroshima University*, Series A–I, **28**, 89–93.
- Monderer D and Shapley LS (1996) Potential games. *Games and Economic Behavior*,

**14**, 124–143.

Parthasarathy T, Tijs SH and Vrieze OJ (1984) Stochastic games with state independent transitions and separable rewards. In: Selected topics in OR and Mathematical Economics (Eds: G. Hammer and D.Pallaschke) Springer Verlag, Berlin, Lecture Note Series 226, 262–271.

Raghavan TES and A. S. Nowak (1993), A finite step algorithm to find a Nash equilibrium point in stationary strategies for non-zero sum one player control discounted stochastic games, *Mathematical Programming* **59** , 249-259.

Raghavan TES, Fe Evangelista and O.J. Vrieze (1996), Repeated ARAT games, In: Statistics, Probability and Game Theory, Edited by T.S. Ferguson and L.S. Shapley, IMS Lecture Notes Series vol. 30, pp 13-28.

Raghavan TES and F. Thuijsman (1997) Perfect information stochastic games and related classes, *International J. Game Theory*, **26**, (1997), pp. 403-408.

Raghavan TES, Tijs SH and Vrieze OJ (1985) On stochastic games with additive rewards and additive transitions. *Journal of Optimization Theory and Applications*, **47**, 451–464.

Shapley LS (1953) Stochastic games. *Proceedings of the National Academy of Science USA*, **39**, 1095–1100.

Shapley LS (1964) Some topics in two-person games. In: Advances in game theory (Eds. M Dresher, LS Shapley and AW Tucker). Princeton University Press, Princeton NJ, 1–28.